

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2002, 20:58:41 ; Search time 1527.5 seconds
(without alignments)
16168.980 Million cell updates/sec

Title: US-10-025-514-15
Perfect score: 1525
Sequence: 1 tctagaccatggaagacct.....ccagtcaggcctagtgcac 1525

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_Other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query Length | ID | Description |
|------------|-------|-------|--------------|-------------|--------------------|
| 1 | 429 | 28.1 | 2478 | 11 AF130068 | AF130068 Homo sapi |
| 2 | 427.4 | 28.0 | 2571 | 11 AF113676 | AF113676 Homo sapi |
| 3 | 331.2 | 21.7 | 1392 | 11 AK002537 | AK002537 Mus muscu |
| 4 | 318.4 | 20.9 | 1296 | 11 AK004999 | AK004999 Mus muscu |
| 5 | 309.8 | 20.3 | 887 | 14 BQ643710 | BQ643710 AGENCOURT |
| 6 | 305.4 | 20.0 | 907 | 14 BQ648909 | BQ648909 AGENCOURT |

| | | | | | | |
|----|-------|------|------|----|----------|--------------------|
| 7 | 286.8 | 18.8 | 895 | 14 | BQ653587 | BQ653587 AGENCOURT |
| 8 | 285.2 | 18.7 | 982 | 14 | BQ646142 | BQ646142 AGENCOURT |
| 9 | 284.2 | 18.6 | 985 | 14 | BM924019 | BM924019 AGENCOURT |
| 10 | 283.8 | 18.6 | 924 | 14 | BQ958958 | BQ958958 AGENCOURT |
| 11 | 274.8 | 18.0 | 1194 | 14 | BM924813 | BM924813 AGENCOURT |
| 12 | 273.4 | 17.9 | 907 | 14 | BQ650189 | BQ650189 AGENCOURT |
| 13 | 270.4 | 17.7 | 959 | 14 | BQ648524 | BQ648524 AGENCOURT |
| 14 | 268.6 | 17.6 | 891 | 14 | BQ646948 | BQ646948 AGENCOURT |
| 15 | 268 | 17.6 | 796 | 12 | BG567260 | BG567260 602589741 |
| 16 | 266.4 | 17.5 | 833 | 14 | BQ007663 | BQ007663 UI-H-EIO- |
| 17 | 264.8 | 17.4 | 991 | 14 | BQ064738 | BQ064738 AGENCOURT |
| 18 | 263.6 | 17.3 | 686 | 10 | AV649258 | AV649258 AV649258 |
| 19 | 262 | 17.2 | 686 | 12 | BG569467 | BG569467 602589937 |
| 20 | 261 | 17.1 | 917 | 14 | BQ650502 | BQ650502 AGENCOURT |
| 21 | 260.2 | 17.1 | 916 | 14 | BQ644130 | BQ644130 AGENCOURT |
| 22 | 260.2 | 17.1 | 977 | 14 | BQ654015 | BQ654015 AGENCOURT |
| 23 | 259.6 | 17.0 | 740 | 10 | AV653778 | AV653778 AV653778 |
| 24 | 259.6 | 17.0 | 944 | 14 | BQ645533 | BQ645533 AGENCOURT |
| 25 | 259.6 | 17.0 | 1012 | 14 | BQ064473 | BQ064473 AGENCOURT |
| 26 | 256 | 16.8 | 869 | 13 | BI219058 | BI219058 602938673 |
| 27 | 253 | 16.6 | 654 | 12 | BG563537 | BG563537 602581944 |
| 28 | 251.4 | 16.5 | 884 | 12 | BG336548 | BG336548 602564956 |
| 29 | 251.2 | 16.5 | 917 | 14 | BQ650698 | BQ650698 AGENCOURT |
| 30 | 250.6 | 16.4 | 746 | 12 | BG566873 | BG566873 602589565 |
| 31 | 250.2 | 16.4 | 745 | 9 | AI114643 | AI114643 HA1237 Hu |
| 32 | 249.8 | 16.4 | 1044 | 9 | AI528087 | AI528087 u193c09.Y |
| 33 | 249.6 | 16.4 | 813 | 13 | BI759000 | BI759000 603042479 |
| 34 | 249.4 | 16.4 | 764 | 12 | BF384318 | BF384318 602046666 |
| 35 | 249 | 16.3 | 731 | 10 | AV649144 | AV649144 AV649144 |
| 36 | 249 | 16.3 | 813 | 10 | BQ182052 | BQ182052 UI-H-EUO- |
| 37 | 246 | 16.1 | 804 | 12 | BG618951 | BG618951 602646361 |
| 38 | 244.4 | 16.0 | 780 | 9 | AI110706 | AI110706 HA0145 Hu |
| 39 | 244.2 | 16.0 | 992 | 13 | BI246204 | BI246204 602958716 |
| 40 | 243.8 | 16.0 | 752 | 12 | BG618001 | BG618001 602644877 |
| 41 | 243.2 | 15.9 | 674 | 14 | BM724546 | BM724546 UI-E-E01- |
| 42 | 243.2 | 15.9 | 759 | 12 | BG533293 | BG533293 601860666 |
| 43 | 243 | 15.9 | 749 | 9 | AI526727 | AI526727 u142b09.Y |
| 44 | 242 | 15.9 | 629 | 12 | BG545257 | BG545257 602572467 |
| 45 | 241.2 | 15.8 | 780 | 12 | BG617594 | BG617594 602615086 |

ALIGNMENTS

| | | | | | | |
|------------|--|---------------|---------|------|---------------|-----------------|
| RESULT 1 | AF130068 | AF130068 | 2478 bp | mrna | linear | HTC 08-MAY-2001 |
| LOCUS | Homo sapiens | clone FLB8226 | PRO2209 | mrna | complete cds. | |
| DEFINITION | AF130068 | | | | | |
| ACCESSION | AF130068 | | | | | |
| VERSION | AF130068.1 | GI:11493442 | | | | |
| KEYWORDS | HTC. | | | | | |
| SOURCE | Homo sapiens. | | | | | |
| ORGANISM | Homo sapiens | | | | | |
| REFERENCE | 1 (bases 1 to 2478) | | | | | |
| AUTHORS | Zhang, C., Yu, Y., Zhang, S., Wei, H., Bi, J., Zhou, G., Dong, C., Zai, Y., Xu, W., Gao, F., Liu, M. and He, F. | | | | | |
| TITLE | Functional prediction of the coding sequences of 75 new genes deduced by analysis of cDNA clones from human fetal liver | | | | | |
| JOURNAL | Unpublished | | | | | |
| REFERENCE | 2 (bases 1 to 2478) | | | | | |
| AUTHORS | Zhang, C., Yu, Y., Zhang, S., Wei, H., Bi, J., Zhou, G., Dong, C., Zai, Y., Xu, W., Gao, F., Liu, M. and He, F. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (23-FEB-1999), Department of Experimental Hematology, Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing, Beijing 100850, P. R. China | | | | | |
| FEATURES | Location/Qualifiers | | | | | |
| source | 1..2478 | | | | | |
| | /organism="Homo sapiens" | | | | | |
| | /db_xref="taxon:9606" | | | | | |
| | /clone="FLB8226" | | | | | |

| | | |
|------------|---|---|
| CDS | /tissue_type="liver" /dev_stage="fetus" 1082..2338 /note="predicted protein of HQ2209" /codon_start=1 /evidence="not_experimental" /product="PRO2209" /protein_id="AA035496.1" /db_xref="GI:11493443" /translation="PPSVSWGILLLAGLCLVPSVLAEDPQGAQAQKTDTSIHDDH PTNKLTPLAEFAFSLROLAHNSNTNIFSPVSTAFAMLSLGTADTHDEILE GLMNFNTEPEAQLIEHGQELRLTNDPSQLQTLTNGLLTSEGSLKLVDFKLEK LYHSEATVFGDGEAKOINDVVEKTKGMFNIOHKLLSSWVLLMVLGNATAIF ERPEFYKDTFEEDFHVDDVTIVKPKMRKLGWNIHKLSDVLDKSVLGLGATIF FLPDEKRLHNEIETHLIIHKFLENEDRRSRASHLEKLSITGTGYDLKSVLGLGATIF VFNEDKLSGVTEEAAPLAKVAHVAVLIDIEKTEAAGAMFLEAPMSIPPEVKFNK PFVFLMEQNTKSPFLMGKVNPTOK" | |
| BASE COUNT | 650 a | 627 c 597 g 604 t |
| ORIGIN | | |
| | Query Match | 28.1%; Score 429; DB 11; Length 2478; |
| | Best local Similarity | 60.2%; Pred. No. 6.6e-101; |
| | Matches 711; Conservative | 0; Mismatches 470; Indels 0; Gaps 0; |
| QY | 12 | GAAGACCTCAAGCGAGCGCGCTCAAAAACCGACACCGAGTCATCAGACCAAGACCAT 71 |
| DB | 1154 | GAGGATCCCGAGGAGATGCTGCCAAGACAGATACATCCACCATGATCAGGATCAC 1213 |
| QY | 72 | CCGACTTTTAATAAANTACTCCAANTTAGCCGAATTTGCTTTTCTTGATAGACAA 131 |
| DB | 1214 | CCAACTTTCAACAGATACACCCCAACCTGGCTGAGTTCGCTTCAGCCTATACGCCAG 1273 |
| QY | 132 | TTAGTCATCAAGTAATTTCTACTAAATTTTTTTTAGTCCTGTTTCTATTGCCACTGCT 191 |
| DB | 1274 | CTGGCACACCGAGTCCACAGACCAATATCTTCTCTCCAGTGAGCATCGCTACAGCC 1333 |
| QY | 192 | TTGCGCATGTTGAGTTTAGTGACTTAAGCCGATACCCATGACGAGATTTTAGAAGTTTA 251 |
| DB | 1334 | TTTGCAATGCTCTCCCTGGGACCAAGGTGACACTCAGATGAATCTCTGGAGGCCGTG 1393 |
| QY | 252 | AACTTTAAATTCACCGAATCCCAAGAAGCCCAATTCACGAGGGTTTTCAGAGTTGTTG 311 |
| DB | 1394 | AATTTCAACCTCAGCGAGATTCGAGGCTCAGATCCATGAAGGCTTCACGAACTCCTC 1453 |
| QY | 312 | AGAACTTTGAATCAACCTGATTCATTTGGAATTAACACTACTGGTAAACGGTTATTTTG 371 |
| DB | 1454 | CGTACCCTCAACACGCCACAGACGACGCTCCAGCTGACACCGGCAATGGCCTGTTCTTC 1513 |
| QY | 372 | TCGTGAAGGTTTAAATTTGGTTGACAAATTCCTAGAAGACGTCGAAGAACTATATCATAGT 431 |
| DB | 1514 | AGCGAGGGCTGAACCTAGTGGATAAGTTTGTGAGGATGTTAAAAGTTGTACCACTCA 1573 |
| QY | 432 | GAGGCTTTTACCGTTTAATTTGGTGATACTGAGGAAGCTAAAAGCAAATTAATGATTAAT 491 |
| DB | 1574 | GAAAGCCTTCACCTGCAACTTCGGGCACACCGAAGAGGCCAAGAACAGATCAACGATTAC 1633 |
| QY | 492 | GTTGACAAAGGACCCAGGGTAAGTCTGTACCTACTTAAAGAAATTAGATCGTGATACC 551 |
| DB | 1634 | GTGGAGAGGGGTACTCAAGGGAAATTTGGATTTGGTTCAGAGAGCTTGACAGAGACACA 1693 |
| QY | 552 | GTCTTCGCACTAGTTAACTATATTTTTTCAAGGGTAAGTGGGAAGCTCCTTTCCAGGTT 611 |
| DB | 1694 | GTTTTTGCTCTGTTGATTTACATCTCTTAAAGSCAAATGGGAGAGACCCCTTTGAAGTC 1753 |
| QY | 612 | AAAGATACTGAGAGAGAGATTTTCATGTTGATCAAGTTACTACTGTCAAAAGTCCAATG 671 |
| DB | 1754 | AAGCACACCGGAGAGAGACTTCCACCTGACACGAGTGCACCCGTGAAGGTGCCTATG 1813 |
| QY | 672 | ATCAANAAGACTGGGTATGTTCAATATTCACATTTGCAAAAAAATTAAGTTCTTCTGCTTAA 731 |
| DB | 1814 | ATGAGCGGTTTAGGCATGTTTAACATCAGCACTGTGAAGAGCTGTCCACCTGGGTGCTG 1873 |
| QY | 732 | TTAATGAAGTATTTAGGTAAACGCTACTGCTATTTTTTTTTTTTATTTTACACAGCAAGGTAGCTTT 791 |

/translation="MPSSVSMGILLLAGLCLVPSVLAEDPQDAQAQKTDTRSHHDQDH
PTFNKLTIPNLAEPSPYQRLAHSNNTNIEFSPVSIATAFAMLSLGTADTHDEILE
GLNFPEIIPAFQIHGEOELLRTLNOPDSOLOLTTNGLELSGLKDFLEDVKK
LYHSEAFVNFDEFEAKKOINDYVEKGTQSKIYDLVYKELDRDVFALVNIIFKRW
ERPPEKVTEDBEDHVDQVTVKPMARLGMFNHIOHCKLUSWVLLMKYLGUNATIF
FLPDGKLOHLENEHDIITIKFLENERDRRSASLHPLKLSITGYDLSVLOLGITK
VFSNGADLSGVTEAPLEAPLKLKSAVHKAVLTIDEKGTAAAGAFLEAIPMSIPPEVKFNK
PFVFLMIQNTKSPFLMGKVVNPQK*
BASE COUNT 800 a 638 c 556 g 577 t

| Query Match | 28.0% | Score 427.4 | DB 11 | Length 2571 |
|-----------------------|--|--------------------|----------|-------------|
| Best Local Similarity | 60.1% | Pred. No. 1.7e-100 | | |
| Matches 710 | Conservative 0 | Mismatches 471 | Indels 0 | Gaps 0 |
| QY 12 | GAAGACCCCTCAAGGCGACCGCTCAAAACCGACACACGACGATCATCAGCACCAAGACCAT 71 | | | |
| DB | 1181 GAGGATCCCGGAGATGCTCCCGAAGACAGATACATCCACCATGATCAGGATCAC 1240 | | | |
| QY 72 | CCGACTTTTAAATAAATTAATCTCAAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA 131 | | | |
| DB | 1241 CCAACCTTCAACAAGATCACCCCAACCTGGCTGAGTTCCGCTTTCAGCCCATACCGCCAG 1300 | | | |
| QY 132 | TTAGCTCATCAAGTAATTTCTACTAACATTTTATTTAGTCTGTTCTTATTCGCTACTGCT 191 | | | |
| DB | 1301 CTGGCACACAGTCCCAACAGCACAATATCTTCTTCTCCAGTGAAGTCTGCTACAGCC 1360 | | | |
| QY 192 | TTCCGCATGTTGAGTTAGTACTAAGCCGATACCCATGACGAGATTTTGAAGGTTTA 251 | | | |
| DB | 1361 TTTCGCACTCTCCCTGGGGACCAAGGCTGACACTCAGATGAATCTTGGAGGCGCTG 1420 | | | |
| QY 252 | AACCTTTAATTTGACCGAATCCAGAACCCCAATTTACGAGGGTTTTCAGAGAGTTGTTG 311 | | | |
| DB | 1421 AATTTCAACCTCAGGAGATTCGGAGGCTCAGATCCATGAAGGTTTCAGGAATCCTC 1480 | | | |
| QY 312 | AGAACCTTGAATCAACCTGATTTCTCAATTCGAATTAATCTGTTGTAACGGTTTATTTT 371 | | | |
| DB | 1481 CGTACCTTCAACAGCCAGACAGCCAGCTCCAGCTGACCAACCGCAATGGCTGTTCTC 1540 | | | |
| QY 372 | TCGTAAGGTTTAAATTTGTTGATATCTGAGGAGCTTAAAGCAAAATTAATGATTAT 491 | | | |
| DB | 1601 GAAGCCTTCACTGTCAACTTCGGGACACCGGAGGCGCAAGAACATCAACGATTAC 1660 | | | |
| QY 492 | GTGGAAGAGGACCCAGGTTAGATCGTTGACCTAGTTTAAGAAATTAAGATCGTGATACC 551 | | | |
| DB | 1661 GTGGAGAAGGGTACTCAAGGGGAAATTTGATTTTGGTCAAGGAGCTTGACAGAGACACA 1720 | | | |
| QY 552 | GTCTTCGCCTAGTTTAACTATATTTTTCAGGAGTAAGTGGGAGCGTCTTTCGAGGTT 611 | | | |
| DB | 1721 GTTTTGTCTGTTGAATTAATCTTCTTAAAGCAATGGGAGAGACCTTTTGAAGTC 1780 | | | |
| QY 612 | AAAGTACTGAAGAGAGATTTTTCATGTTGATCAAGTTACTACTGTCAAAAGTTCCCAATG 671 | | | |
| DB | 1781 AAGGACCCGAGGAGAGGACTTCCACGTGGACAGGTTGACCAACCGTGAAGTGCCATG 1840 | | | |
| QY 672 | ATGAAGAGACTGGGTATGTTCAATATTCACATTCGCAAAATTAAGTTCTTTGGTCTTTA 731 | | | |
| DB | 1841 ATGAAGCGTTTAGGCGATGTTTAAACATCCAGCACTGTAAGAAGCTGTCCAGCTGGGTG 1900 | | | |
| QY 732 | TTAATGAAGTATTTAGTTAAGCTACTGCTATTTTTCAGGAGGAGGAGGTAAGCTT 791 | | | |
| DB | 1901 CTGATGAATACCTGGGCAATGCCACCCCATCTTCTTCTGCTGATGAGGGGAACTA 1960 | | | |
| QY 792 | CAACATTTAGAGATAGTTGACTCATGACATTAATTTACTAAATTTTATGAGAAGCAGGAT 851 | | | |
| DB | 1961 CAGCACCTGGAAATGAACCTCACCCAGCATATCATCAACAGTTCTCTGGAAATGAAGAC 2020 | | | |
| QY 852 | CGTCGTAGCGCTTCTCTCGACCTGCGCAAGTTAGTATCATCCGGTACTTACGACTTAAA 911 | | | |

| | | | | | |
|------------|--|--|--|--|--|
| Db 2021 | AGAAGTCTGCCAGCTTACATTTTACCACAACTGTCATTACTGGAACCTATGATCTCAAG 2080 | | | | |
| QY 912 | TCTGTTTATAGCCAGTTAGTATTACCAAGTTTTTTTCTTAAGGTCGCCGATTTGAGTGGT 971 | | | | |
| Db 2081 | AGCGTCTCGGTCAACTAGGCGATCACTAAGGCTTCTAGCAATGGGCGTACCTCTCCGGG 2140 | | | | |
| QY 972 | GTTTACTGAAGAAGCTCCATTTAAATTTAGTAAAGCTGTTTCAAAAGCGCTTTAACTATT 1031 | | | | |
| Db 2141 | GTACAGAGGAGGACCCCTGAAGCTCTCCNAGCCGCTGCATAAGGCTGTGCTGACCATC 2200 | | | | |
| QY 1032 | GATGAAAGAGGTACCGAGGCGCGCGCTATGTTCTTGGAGGCTATTTCAATGAGGATTT 1091 | | | | |
| Db 2201 | GACGAGAAAGGACTGAAGCTGCTGGGCGCATGTTTTTAGAGGCCATACCATCTCTATC 2260 | | | | |
| QY 1092 | CCACCAGAGTTAAATTTAATAAACCATTCGTTTTTCTGATGATCGAGCAGACACATAA 1151 | | | | |
| Db 2261 | CCCCCGAGGTCAAGTTTCAACAAACCCCTTTGCTTCTTAATGATGAACAAAATACCAAG 2320 | | | | |
| QY 1152 | AGCCATTTGTTATGGTGAAGTTTGTCAACCCAACTCAGAA 1192 | | | | |
| Db 2321 | TCTCCCTCTTCATCGGAGAAAGTGTGAATCCCAACCCAAA 2361 | | | | |
| RESULT 3 | | | | | |
| AK002537 | | | | | |
| LOCUS | AK002537 1392 bp mRNA linear HTC 19-JAN-2002 | | | | |
| DEFINITION | Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610011G14:serine protease inhibitor 1-4, full insert sequence. | | | | |
| ACCESSION | AK002537 | | | | |
| VERSION | AK002537.1 GI:12832592 | | | | |
| KEYWORDS | HTC; CAP trapper. | | | | |
| SOURCE | Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone:0610011G14. | | | | |
| ORGANISM | Mus musculus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | | | |
| REFERENCE | 1 | | | | |
| AUTHORS | Carninci, P. and Hayashizaki, Y. | | | | |
| TITLE | High-efficiency full-length cDNA cloning | | | | |
| JOURNAL | Meth. Enzymol. 303, 19-44 (1999) | | | | |
| MEDLINE | 99279253 | | | | |
| PUBMED | 10349636 | | | | |
| REFERENCE | 2 | | | | |
| AUTHORS | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. | | | | |
| TITLE | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes | | | | |
| JOURNAL | Genome Res. 10 (10), 1617-1630 (2000) | | | | |
| MEDLINE | 20493374 | | | | |
| PUBMED | 11042159 | | | | |
| REFERENCE | 3 | | | | |
| AUTHORS | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, I., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Takeda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. | | | | |
| TITLE | RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipipette sequencer | | | | |
| JOURNAL | Genome Res. 10 (11), 1757-1771 (2000) | | | | |
| MEDLINE | 20530913 | | | | |
| PUBMED | 11076861 | | | | |
| REFERENCE | 4 | | | | |
| AUTHORS | Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., | | | | |


```

/gene="Spill-4"
/Note="data source:MGI, source key:MGI:891968,
evidence:ISS
putative
serine protease inhibitor 1-4"
/codon_start=1
/protein_id="BAB23733.1"
/db_xref="GI:12836612"
/db_xref="MGI:891968"
/translation="MTPSISWCLLLGLLGLCLVPSFLAEDVQETDTSKQDSPAHEI
ANLGFATSLRYELVHOSNTNIFSPVSIATAFAMLSLGSKGDTHQILLEGQFNL
TQSEADHNSFOHLLQTLNRDSELOLSTGNLEVNNDLKLVERFLKAEKNHYOAEV
FSYFADIESEAEKKVINDFVQKQIVKLEQDQFVYLANVLLKQWKPFDP
ENTKQAEFHVDESTTVKVPMTLSGLMDLVHSCSTLSSWVLLMDYAGNATFAPFLLEDG
KMOHLEOTLNKELISKFLNRRRLRAQIHIPRLSISGNYNLETLSPGILTRIFNSGA
DLSGITTEENAPLKLSQAVHKAVITIDETGTEAAATVLLQGGFLSPPLHFNRPFLFI
IFEHSQSPLFVGVKVDPTK"
BASE COUNT      339 a      370 c      300 g      287 t
ORIGIN

Query Match      20.9%; Score 318.4; DB 11; Length 1296;
Best Local Similarity 55.6%; Pred. No. 4.1e-72;
Matches 654; Conservative 0; Mismatches 516; Indels 6; Gaps 2;

QY 20 TCAAGGCGACCGCGCTCAAAAACCGACACAGCATCATCAGCAACCAACCATCGACTTT 79
DB 113 TCTGCTGAGGATGTTAGGAGACAGACACCTCCAGAGGATCATGCC---CCAGCCTC 169
QY 80 TAATAAATTTACTCAATTTAGCGGATTTGCTTTTCTTTGATATAGACAATTAGCTCA 139
DB 170 CCATGAGATTTGCTACAAACCTGGGAGACTTTGCCATCAGTCTATACCGGGAGCTTGCAT 229
QY 140 TCAAGATTAATTTACTAATTTTCTTTAGTCTGTTTCTTTATTCCTCCTGTTTTCGCCAT 199
DB 230 TCGTGCAACACTTCCACATCTTCTTCTCCAGTGAGCATTTGCCACAGCCTTTGCTAT 289
QY 200 GTTGAGTTTGTAGTACTAAGCGGATACCCATGACGAGATTTTGAAGGTTTAACTTTAA 259
DB 290 GCTCTCCCTAGGAGGAGGAGGTTGACATTCACACGAGCATCTCAGAGGCGCTGCGAGTTCAA 349
QY 260 TTTGACCGAATCCAGAGGCGCAATTCACGAGGTTTCAAGAGTTGTTGAGAACTTT 319
DB 350 CCTCACAACAACATCGGAGGCTGACATCCCAATTCCTTCCAACACTCTCCCAACCT 409
QY 320 GAATCAACTGATTTCTCAATTTCAATTTAACTACTGTTTAACTGTTTAACTGTTGTAAGG 379
DB 410 CAACAGACACAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 469
QY 380 TTTAAATTTGTTGACAAATTCCTAGAGAGAGTCAAGAACTATATATAGTGTGAGGCTTT 439
DB 470 CCTGAGCTGGTGGAGAGTTTCTGGAAGAGGCGCAAGAACCAATTTATGAGGAGAGTCTT 529
QY 440 TACCGTTAATTTGGTGATGATGAGAGAGTCAAGAACTAATTAATGATTTATGTTGAGAA 499
DB 530 CTCTGTCACTTTGCGAGATCAGAGAGGCGCAAGAACTAATTAATGATTTATGTTGAGAA 589
QY 500 AGGACCCAGGCTAAGTCTGACCTAGTCTAAGAACTTAAAGAACTTAAAGTGTGATGATGCTTGC 559
DB 590 GGAACCCAGGAAAGATAGTTGAGGCTGTGAAAGAACTTAAAGAACTTAAAGTGTGATGCTTGC 649
QY 560 ACTAGTTAATTTATTTTCAAGGTAAGTGGGACCTCTCTTTCAGGTTTAAAGATAT 619
DB 650 CTTGGCAAAATTTACATCTCTTTAAAGCAATTTGGAAGAGCAATTCGATCTTCAAGACAC 709
QY 620 TGAAGAGGAAGATTTTCATGTTGATCAAGTTACTACTGTCAAGTTTCAAGTTTCAAGTGAAG 679
DB 710 TAAGCAAGCTGAGTTTCCAGTGTGACGAGTCCACCGGTGAAGTGTCCCATGATGACCT 769
QY 680 ACTGGGTATGTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTGA 739
DB 770 CTCGGGCTGCTTGCAGTGTGACCATTTGCAGCAGCTCTCCAGTGTGGTGTGCTGTGAGGA 829
QY 740 GTATTTAGGTAACGCTACTGCTATTTTTTTTACCAGAGGAGGTAAGCTTCAACATTT 799

```

```

DB 830 TTACGCGGCAACGCCACTGCTGCTTCTCCTCCTCCCGATGATGGAGATGCGAGCATCT 889
QY 800 AGAGAATGAGTTGACTCATGACATTAATTAATAAATTTTATAGAACAGGAGGATGCTGCTAG 859
DB 890 GGAGCAAACTCTCAACAAGAGGACTCATCTTAAGTCTTCTGCTAAACAGGCGCAGAAGTT 949
QY 860 CGCTTCTCTGCACTGCGCCAAAGTTAAGTATCACCGGTACTTACGACTTAAATCTGTTTT 919
DB 950 AGCCAGATCCATATCCCGAGACTGCTCATCTCTGGAACCTATAACTTTGGAGACACTCAT 1009
QY 920 AGGCCAGTAGTATTACAAAGTTTTCCTTACCGTCCCGATTTGAGTGTGTTACTGA 979
DB 1010 GAGTCCACTGGGCATCATCCCGGATCTTCAACAGTGGGGCTGACCTCTCCGGAATCAGAGA 1069
QY 980 AG---AAGCTCATTAATAATTCAGTAAAGCTTTCACAAAGCCGCTTCTTAACATTTATGATGA 1036
DB 1070 GGAGAATGCTCCCTGAGAGCTCAGCCAGGCTGTGCATTAAGCCGCTGCTGACCATCATGA 1129
QY 1037 AAAGGTACCGAGCGCGCGCTATGTTCTGGAAGCTATTCCAATGAGCATTTCCAC 1096
DB 1130 GACAGAACAGAGCTGCGAGCAGCTACAGCTTACAAAGCGGCTTTTGTCTATGCCCTC 1189
QY 1097 AGAAGTTAAATTTAATAAACATTTGTTTCTGATGATCGAGCAGACACACTTAAAGGCC 1156
DB 1190 TATCTTGCACCTTCAACCGCCTTTCTTCTTCAATAATTTTGAAGAACACTCTCAGAGGCC 1249
QY 1157 ATGTTTATGGGTAAAGTTGTCAACCCAACTCAGAA 1192
DB 1250 CCTCTTTGTGGGAAAAGTGTGATATCCACACATAA 1285

RESULT 5
BQ643710      887 bp      mRNA      linear      EST 15-JUL-2002
LOCUS      AGENCOURT_8342217 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268225
5', mRNA sequence.
BQ643710
VERSION      BQ643710.1 GI:21767882
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 887)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2439 row: n column: 02
High quality sequence stop: 731.
FEATURES
source
1..887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6268225"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)

```

using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

| BASE COUNT | 242 a | 224 c | 232 g | 187 t | 2 others |
|---------------------------|--|--|-------|-------|----------|
| ORIGIN | | | | | |
| Query Match | 20.3%; Score 309.8; DB 14; Length 887; | | | | |
| Best Local Similarity | 60.0%; pred. No. 6.5e-70; | | | | |
| Matches 532; Conservative | 0; Mismatches 354; Indels 1; Gaps 1; | | | | |
| Qy | 195 | GCCATGTTGAGCTTAGGTACTAAAGCCGATACCCATGACGAGATTTTGAAGGTTTAAAC | 254 | | |
| Db | 1 | GCATGCTCTCCCTGGGGACCAAGGCTCACACTCACGATGAATCTCGAGGGCTGAAT | 60 | | |
| Qy | 255 | TTTAATTTGACCGAAATCCCAAGAGCCCAAAATTCACGAGGGTTTTCAAGAGTTGTTGAGA | 314 | | |
| Db | 61 | TTCAACCTCACGGAGATTCGGGAGGCTCAGATCCATGAAGGCTTCCAGGAACCTCTCCGT | 120 | | |
| Qy | 315 | ACTTTGAATCAACCTGATCTCAATTTGCAATTAACCTACTGGTAGCGGTTTATTTTGTCT | 374 | | |
| Db | 121 | ACCTCTAACACGACAGACAGCAGCTCAGCTGACACCGCAATGGGCTGTTCTCTCAGC | 180 | | |
| Qy | 375 | GAAGGTTTTAAATTTGGTTTGACAAATTCCTACAAGAGCTCAAGAACTATATCATAGTGAG | 434 | | |
| Db | 181 | GAGGCGCTGAAGCTAGTGATTAAGTTTTTGGAGATGTTTAAAGTGTCTACCACTCAGAA | 240 | | |
| Qy | 435 | GCTTTTACCGTTAATTTTGGTGATGACGAGGCTAAAAAGCAAAATTAATGATTATGTT | 494 | | |
| Db | 241 | GCTTTCAGTGTCACTTCGGGGACACCGAAGAGGCCAAGAAACAGATCAACGATTACGTG | 300 | | |
| Qy | 495 | GAGAAAGCACCCAGGGTAAGATCGTTGACCTAGTTAAAGAAATTAGATCGTGATACCGTC | 554 | | |
| Db | 301 | GAGAAGGTACTCAAGGGAATTTGTGGATTTGGTCAAGAGCTTTGACAGAGACACAGTT | 360 | | |
| Qy | 555 | TTCCGACTAGTTACTATATTTTTTCAAGGGTAAGTGGGACGCTCTTTCCAGGTTAAA | 614 | | |
| Db | 361 | TTTGCTCTGGTGAATATACATCTCTTTAAAGGCAAAATGGGAGAGACCCCTTTGAAGTCAAG | 420 | | |
| Qy | 615 | GATACTGAAGAGGAAGATTTTCATGTTGATCAAGTTACTACTGTCAAAAGTTCCCAATGATG | 674 | | |
| Db | 421 | GACACCGAGGAAGAGGACTTCCAGCTGGACCAAGTGACCACCGTGAGGTGCCCTATGATG | 480 | | |
| Qy | 675 | AAAAGACTGGGTATGTTCAATATTCAATTTGCAAAATTAAGTTCTTGGGCTTATTTA | 734 | | |
| Db | 481 | AAGCGTTTTAGCGATGTTTAAGATCCAGCACTGTGAAGAGCTGTCCAGCTGGGTGCTGCTG | 540 | | |
| Qy | 735 | ATGAAGTATTTAGTAAGGCTACGCTACTGCTATTTTTTTTACCAGAGGAGTTAACTTCAA | 794 | | |
| Db | 541 | ATGAATATCTGGGCAATGCCACCGCATCTTCTCTGCTGTATGAGGGGAAACTACAG | 600 | | |
| Qy | 795 | CATTTAGAGATAGTTTGACTCATGACATTATTACTAAATTTTTTAGAGAACGAGGATCGT | 854 | | |
| Db | 601 | CACCTGGAAATGAATCAACCACGATATCATCAAGTTCTCTGGAAATGAAGACAGA | 660 | | |
| Qy | 855 | CGTAGCGCTTCTGCACTGCCAAAGATTAGTATACCGGTACTTACGACTTAAATCT | 914 | | |
| Db | 661 | AGGTCTGCCAGCTTACATTTTACCACCAATGTGCCATTACTGGAACTATGATCTGAAGAGC | 720 | | |
| Qy | 915 | TTTTTAGCGCAGTTAGGTATTACCAAAAGTTTTTCTTAAGGGTGGCGATTTTCAGGTGGTGT | 974 | | |
| Db | 721 | GTCTGNGTCAACTGGGCATCATTAAAGTCTTCAGCAATGGGGCTGACCTCTCCGGNGTC | 780 | | |
| Qy | 975 | ACTGAAGAGCTCATTTAAATTTGAGTTAAAGCTGTTCCAAAGCCGCTTTAACTATTGAT | 1034 | | |
| Db | 781 | ACAGAGGAGGACCCCTGAAGCTCTCCAAGGC-GTGCAATAGGCTGTGCTGACCACTGAC | 839 | | |
| Qy | 1035 | GAAGAGGTTCCGAGGCGCGGGCGCTATGTTCTCTGGAACTATTCC | 1081 | | |
| Db | 840 | GAGAAAGGACTGAAGCTGTGGGGCCATGTTTTTAGAGGCCATATCC | 886 | | |

```
QY 577 TTTTCAAGGTAAGTGGGAACGCTCTTTCAGAGTTTAAAGATACTGAAGAGGAAGATTTTC 636
Db 375 TCITTAAGGCAATGGAGAGACCCTTTGAAGTCAAGGACACCGAGGAGAGGACTTCC 434
QY 637 ATGTTGATCAAGTTACTGTCGCAAGTTCCTCAATGATGAAGAAACACATGGGTATGTTCAATA 696
Db 435 ACGTGGACCAAGTGACCAACGCTGAAGGTGCTATGATGAAGCGTTTAGGCATGTTTAAACA 494
QY 697 TTCAACATTCGCAAAATTAAGTTCCTGGGCTCTATTAAATGAAGTATTTAGGTAACGCTA 756
Db 495 TCCAGCACTGTAAGAAGCTGTCCAGCTGGGTGCTGCTGATGAATACCTGGGCAATGCCA 554
QY 757 CTGCTATTTTTTTTTTACCAGACGAAGCTAAGCTTCAACATTTAGAGAATGAGTTGACTC 816
Db 555 CGGCATCTTCTCTGCTGATGAGGGAACACTACAGCACCTGGAAATGAATCACTCACCC 614
QY 817 ATGACATATTACTAAATTTTTAGAGAACGAGGATCGTCTAGCGCTTCTCTGCACCTGC 876
Db 615 ACATATATCATCACAAGTTCTCTGAAATGAAGACAGAAGGCTGTGCCAGCTTACATTAC 674
QY 877 CAAAGTTAAGTATCACCGGTACTTTACGACTTAAATCTGTTTTAGGCCAGTTAGGTATTA 936
Db 675 CCAACTGTCTACTTGGAACTATGATCTGAAGACGCTCCTGGTCACTGGGCATCA 734
QY 937 CCAAGTTTTTCTAAGGTTCCGATTTAGTGGTGTACTGAAGAGGCTCCATTAAT 996
Db 735 CTAAGTCTTCAGCAATGGGCTGACCTCTCGGGGTACAGAGGAGGACCCCTGAAGC 794
QY 997 TGAGTAAAGCTGTTCAAAAGCCGCTCTTAAGTATTGATGAAGAGGTTACCGAGCGCCG 1056
Db 795 TCTCCAGGCG-GTGCAATAAGGCTGTGTCACATCGACGAGAGGAGGACTGAAGCTGCTG 853
QY 1057 GCGCTATGTTCTCGAAGCTATTCCATGAGCATTCACACAGAAAGTT 1103
Db 854 GGGCAATGTTTGAAGGCAATACCATGCTAATCCCCCGAGGT 900

RESULT 7
B0653587 895 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8349591 NIH_MGC_100 Homo sapiens cdna clone IMAGE:6284550
DEFINITION 5', mRNA sequence.
ACCESSION B0653587
VERSION B0653587.1 GI:21777759
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 895)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cdna Library Preparation: Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LICM2482 row: f column: 07
High quality sequence stop: 708.
Location/Qualifiers
1..895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6284550"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: xhoI; Site_2:
```

RESULT 8

EcORI: cdna made by oligo-dT priming. Directionally cloned into EcORI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 227 a 248 c 230 g 190 t
ORIGIN

Query Match 18.8%; Score 286.8; DB 14; Length 895;
Best Local Similarity 60.1%; Pred. No. 6.6e-64;
Matches 477; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

```
QY 12 GAAGACCCCTCAAGGGGACGCGCTCAAAAACCGACACACGATCATCACGACCAAGACCAT 71
Db 95 GAGGATCCCGGAGAGATGCTGCCAGAGAAGACAGATATCCACCATGATCAGGATCAC 154
QY 72 CGGACTTTTAATAAATTTACTCCAAATTTAGCCGAATTTGCTTTTCTTGTATAGACAA 131
Db 155 CCAACCTTCAACAAGATCACCCCAACCTGGCTGAGTTGCGCTTCAGCCTATACCGCCAG 214
QY 132 TTAGCTCATCAAAAGTAATTTCTACTAACATTTTTTTTAGTCTCTGTTTTCTATTGCCACTG 191
Db 215 CTGGCACCGATCCACAGACCAATATCTTCTCTCCCAAGTGAAGCATCGCTACAGCC 274
QY 192 TTCGCCATTTGAGTTTAGGTACTAAAGCCGATACCCATGACGAGATTTTGAAGAGTTTA 251
Db 275 TTTGCAATGCTCTCCCTGGGGACCAAGGCTGACACTCAGCATGAAATCTCGAGGGCGCTG 334
QY 252 AACTTTAAATTTGACCGAAATCCAGNAGCCCAATTCACGAGGTTTTCACAGAGTTGTG 311
Db 335 AATTTCAACCTCACGGAGATTCGGAGGCTCAGATCCATGAAGGCTTCAGGAACTCTC 394
QY 312 AGAAGCTTTGAATCAACCTGATTTCTCAATTCGAATTAACCTACTGTAACCGTTTATTTTG 371
Db 395 CGTACCTCAACGACGACAGCCAGCTCCAGCTGACCACCGCAATGGCTGTTCTCTC 454
QY 372 TCTGAAGGTTTAAATTTGGTTGACAAATTCCTAGAGAGCTCAAGNAACTATATCATAGT 431
Db 455 AGCGAGGGCTGAAGCTAGTGGATAAGTTTTTGGAGGATGTTAAAAAGTTGTACCACTCA 514
QY 432 GAGGCTTTTACCGTTAATTTTGGTGATGATCTGAGGAAGCTAAAAAGCAAAATTAATGATT 491
Db 515 GAAGCCTTCACTGCTCACTTCGCGGACACCGAGAGGCCAAGAAACAGATCAACGATTAC 574
QY 492 GTTGAGAAAGGCCACCGAGGTAAGATCGTTGACCTAGTTTAAAGAAATAGATCGTATACC 551
Db 575 GTGGAGAGGGTACTCAAGGGAAATTTGGATTTTGGTCAAGGAGCTTGACAGAGACACA 634
QY 552 GTCTTCGCCTAGTTAACTATATTTTTTCAAGGGTAAGTGGGAAGCTCTTTCGAGGTT 611
Db 635 GTTTTGTCTCTGGTGAATTACATCTCTTTAAGGCAATGGGAGAGACCCCTTTGAAGTC 694
QY 612 AAAGATACTGAAGAGGAAGATTTTCATGTTGATCAAGTTACTACTGTCAAAAGTTTCAATG 671
Db 695 AAGCACACCGAGGAGGACTTCCAGCTGGACAGCTGACCACCGGAAGGTGCTCATG 754
QY 672 ATGAAAGACTGGGTATGTTTCAATATTAACATTTGCAAAAATTAAGTTCTTGGGTCTTA 731
Db 755 ATGGAGCGTTTAGGCATGTTTAACTCCAGCACTGTGAAGAAGCTGTCCAGCTGGGCTG 814
QY 732 TTATGAAGTATTTAGGTAACGCTACTGCTATTTTTTTTTTACCAGACGAAGTAAGCTT 791
Db 815 CTGATGAATACCTGGGGCATGCCCGCCATCTTCTTCTGCTGGATCAGGGAACCTA 874
QY 792 CAACATTTAGAGAA 805
Db 875 CAGCACCTGGGAAA 888
```


BQ646142
 LOCUS BQ646142 982 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8492569 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6296341
 5', mRNA sequence.
 ACCESSION BQ646142
 VERSION BQ646142.1 GI:21770314
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 982)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM2504 row: a column: 14
 High quality sequence stop: 647.
 Location/Qualifiers
 1. 982
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6296341"
 /clone_lib="NIH_MGC_100"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GCCACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."
 BASE COUNT 267 a 259 c 238 g 217 t 1 others
 ORIGIN
 Query Match 18.7%; Score 285.2; DB 14; Length 982;
 Best Local Similarity 59.8%; Pred. NO. 1.8e-63;
 Matches 496; Conservative 0; Mismatches 333; Indels 1; Gaps 1;
 QY 303 GAGTTGTTGAGAACTTTGAATCAACCTGATCTCAATTGCAATTAACTACTGGTAACGGT 362
 DB 1 GAACCTCTCGTACCTCAACGAGCAGACGCCAGCTCCAGCTGACCCGCGCAATGGC 60
 QY 363 TTATTTTGTCTGAGAGTTTAAATTTGGTTGACAAATTCCTAGAACGCTCAAGAACTA 422
 DB 61 CTGTTCTCAGGAGGGCTGAAGCTAGTGGATAAGTTTGGAGGATGTTAAAAAGTTG 120
 QY 423 TATCATAGTAGGCTTTTACCGTTAAATTTGGTGATAGTGGAGGCTAAAAGCAATTT 482
 DB 121 TACCACTAGAGGCTTTTCTGCTCAACTTCGGGGACACCGAGAGGCCAAGAAACAGATC 180
 QY 483 AATGATTATGTTGAGAAAGCCAGGCTAAGATCGTTGACCTAGTTAAAGAAATTAGAT 542
 DB 181 AACGATTAGTGGAGAGGGTACTCAAGGAAATTTGGTATTTGGTCAAGAGCTTGAC 240
 QY 543 CTGTATACGCTTCGCACTAGTTAACTATATATTTTTTCAAGGGTAAGTGGGAACGCTCT 602
 DB 241 AGAGACACAGTTTTCCTGCTGCTGTAATTTACATCTCTTTTAAAGGCAATTTGGAGACCC 300
 QY 603 TTCGAGGTTAAAGTACTGAAGAGGAACATTTTCATGTTGATCAAGTTACTACTGTCAA 662
 DB 301 TTGAAGTCAAGGACCCGAGAGAGGACCTTCCACGTGGACAGGTCACCCGCTGAAG 360

QY 663 GTTCCATGATGAAGAAGACTGGGTATGTTCAATATTCAACATTTCAAAAAAATTAAGTTCT 722
 DB 361 GTGCTATGATGAAGCGGTTTAGGATGTTTAACTATCCAGCACTGAAGAAGTCTCCAGC 420
 QY 723 TGGGTCTTATTAAATGAAGTATTAGGTAACGCTACTGCTATTTTATTTTATTTTACCAAGCAA 782
 DB 421 TGGGTGCTGCTGATGAATACCTGGGCAATGCCACCGCATCTTCTTCTGCTGCTGATGAG 480
 QY 783 GGTAAAGCTTCAACATTTTAGAATAGTGTGACTCATGACATTTACTAAATTTTATAGAG 842
 DB 481 GGGAACTACAGCAGCTTGGAAATGAACCTACCCACGATATCATCAAGTTCTCTGCTGAA 540
 QY 843 AACGAGGATCTGCTAGGCTTCTCTGCACCTGCCAAGTTAAGTATCACCGGTACTTAC 902
 DB 541 AATGAAGACAGAGGCTCTGCCAGCTTACATTTTACCAAACTGTCCATTTACTTGGAACTAT 600
 QY 903 GACTTAAATCTGTTTTAGGCGAGTTAGTATTACCAAGT-TTTTCTAACGGTGCGCA 961
 DB 601 GATCTGAAGAGCGTCTCTGGGTCACTGGGATCACTAAGGTCTTTCAGCAATGGGCTGA 660
 QY 962 TTTGAGTGGTGTACTGAAGAAGCTCCATTAATTAAGTAAAGCTGTTTCAAAAGCGT 1021
 DB 661 CCTCTCGGGGTCAAGAGGAGGACCCCTGAAGCTCTCAAGCGCTGCATAAGGCTGT 720
 QY 1022 CTTAACTATTGATGAAGAAGGTACCGAGGCGCGGCGCTATGTTCTGGAAGCTATTCC 1081
 DB 721 GCTGACCATCGACGAGAAAGGACTGAAGCTGTGGGGCCATGTTTGAAGGCCATACC 780
 QY 1082 AATGAGCATTCACACGAGAAGTTAAATTAATTAACCACTTCGTTTTTCTGA 1131
 DB 781 CATGCTATCCCCCCCCGAGGGAGTTCAACCAAACTTTGTCTTCTTA 830
 RESULT 9
 LOCUS BM924019 985 bp mRNA linear EST 12-MAR-2002
 DEFINITION AGENCOURT_6709923 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760183
 5', mRNA sequence.
 ACCESSION BM924019
 VERSION BM924019.1 GI:19374398
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 985)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12806 row: m column: 16
 High quality sequence stop: 707.
 Location/Qualifiers
 1. 985
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5760183"
 /clone_lib="NIH_MGC_116"
 /lab_host="DH10B"
 /note="Organ: pooled colon, kidney, stomach; Vector:
 pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 3 colons, age 26 yo male, 49 yo
 female, 71 yo male colon; 46 yo male kidney, and pool of 2
 stomachs, 62 yo male and 70 yo female. Library is

REFERENCE 1
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES source

oligo-dr primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."

| | | | | | |
|-----------------------|-----------------|--|-----------|-----------------|----------|
| BASE COUNT | 258 a | 270 c | 242 g | 213 t | 2 others |
| ORIGIN | | | | | |
| Query Match | 18.6%; | Score 284.2; | DB 14; | Length 985; | |
| Best Local Similarity | 60.3%; | Pred. No. 3.2e-63; | | | |
| Matches 503; | Conservative 0; | Mismatches 329; | Indels 2; | Gaps 2; | |
| QY | 12 | GAAGACCCCTCAAGGCGCGCTCAAAAAACCGACACCATCATCGACCAAGACAT | 71 | | |
| Db | 92 | GAGGATCCCCAGGAGATGCTGCCAGAGACATATATCCACCATGATCAGGATCAC | 151 | | |
| QY | 72 | CCGACTTTTAATAAATPACTCCAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA | 131 | | |
| Db | 152 | CCAACTTTCAACAAGATCACCCCAACCTGCTGAGTTGCGCTTCAGCCTATACCGGCAG | 211 | | |
| QY | 132 | TTAGCTCATCAAGTAATCTACTAACATTTTTTTTAGTCTCTGTTCTATTTGCCACTGCT | 191 | | |
| Db | 212 | CTGGCACACCATGTCACACGACCAATATCTCTCTCCCGCAGTGAGCATCGCTACAGCC | 271 | | |
| QY | 192 | TTCCGCATGTTGAGTTTAGTACTAAAGCCGATACCCATGACGAGATTTTGAAGGTTTA | 251 | | |
| Db | 272 | TTTGCAATGCTCTCCCTGGGACCAAGCTGACACTCAGATGAATCCCTGGAGGCGCTG | 331 | | |
| QY | 252 | AACTTTAATTTGACCGAATCCAGAGCCCAATTTACAGAGGTTTTCAGAGATTTGTTG | 311 | | |
| Db | 332 | AATTTCAACCTCACGAGATTCGGAGGCTCAGATCCATGAAGGCTTCAGGAACCTCCTC | 391 | | |
| QY | 312 | AGAACTTTGAATCACTGATCTCAATGCAATTAAGTACTGCTAGCGTTTATTTTGG | 371 | | |
| Db | 392 | CGTACCTTCAAGCAGCAGACAGCAGCTCCAGCTGACCCCGCAATGCGCTTCTCCTC | 451 | | |
| QY | 372 | TCTGAAGGTTTAAATTTGGTGACAAATTTCTAGAAGACGTCAGAAACTATATCATAGT | 431 | | |
| Db | 452 | ACCGAGGCGCTCAAGCTAGTGATAAGTTTGGAGGATGTTTAAAGGTTTCTACCACTCA | 511 | | |
| QY | 432 | GAGGCTTTTACGTTAATTTTGGTGATCTAGGAGGCTAAAGCAATTAATGATTAT | 491 | | |
| Db | 512 | GAAGCCTTCTACTGTCAACTTCGGGACACCCGAAGAGGCCCAAGAACATCAACGATTAC | 571 | | |
| QY | 492 | GTGAGAAGGACACCCAGGTAAGATCTGTACCTAGTTTAAAGATTAAGATCGTGATACC | 551 | | |
| Db | 572 | GTGGAGAGGGTACTCAAGGGAATTTGGATTTGGTTCAGAGGCTTGACAGAGACACA | 631 | | |
| QY | 552 | GTCCTTCGACTAGTTAATATATTTTTTCAAGGGTAAGTGGGAACGTCCTTTTCGAGGTT | 611 | | |
| Db | 632 | GTCTTTGCTCTGTTGAATTTACATCTCTTTTAAAGGCAATGGGAGACCCCTTTGAAGTC | 691 | | |
| QY | 612 | AAAGATCTGAAGAGAGAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTTCCAAATG | 671 | | |
| Db | 692 | AAAGGACCCGAGAGAGAGACTTCCACCTGGACAGGTTGACCACTGGAAGCTGCCTATG | 751 | | |
| QY | 672 | ATGAAAGACTCGGTATGTTCAATATTCACCAATTCACCAATTAAGTTCTTTGGTCTTTA | 731 | | |
| Db | 752 | ATGAAGCGTTTAGGATGTTTAAACATCCAGCACTGTAGAGCTGTTCCAGCTGGGTGCTG | 811 | | |
| QY | 732 | TTAATGAAGTATTAGTTAAACGCTACTGCTAT-TTTTTTTTTTACCAAGGATTAAGCT | 790 | | |
| Db | 812 | CTGATGAATACCTGGCAATCCACCCGCACTTCTTCNTGCGCTGATAGGGGAAACT | 871 | | |
| QY | 791 | TCACATTTTAGAGATGAGTTCAGTCAATGACATTAATTAATTTTAGAGAA | 844 | | |
| Db | 872 | ACAGGCACTGGAAAATGAACCTCAC-CAGGATATCATCAAGGTTTCTCTGGAAA | 924 | | |
| RESULT 10 | | | | | |
| BQ958958 | | | | | |
| LOCUS | | | | | |
| | 924 bp | mRNA | linear | EST 21-AUG-2002 | |

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source
1. .924
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6483305"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 251 a 253 c 226 g 193 t 1 others
ORIGIN
Query Match 18.6%; Score 283.8; DB 14; Length 924;
Best Local Similarity 58.4%; Pred. No. 4e-63;
Matches 513; Conservative 0; Mismatches 363; Indels 2; Gaps 1;
QY 84 AAAATTACTCCAAATTTAGCGGAATTTGCTTTCTTTGTATAGACAATTAGCTCATCAA 143
Db 2 AGATACCCCAACCTGGCTGAGTTGCGCTTCAGCCTATACCGCAGCTGGCAGACACAG 61
QY 144 AGTAATTCTACTAACATTTTTTTTAGTCTGTTTCTATTGCGCACTGCTTTGCGCATGTTG 203
Db 62 TCCACAGCAGCAATACTTCTTCTCCCACTGAGCATCGCTACAGCCTTTGCAATGCTC 121
QY 204 AGTTTAGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGTTTAACTTTAATTG 263
Db 122 TCCCTGGGGACCAAGGCTGACACTCAGATGAAATCCTGGAGGCGCTGAATTTCAACCTC 181
QY 264 ACCGAATCCCAAGACCCAAATTCAGAGGGTTTTCAAGAGTTGTTGAGAATTTGAAT 323
Db 182 ACGGAGATTCGGAGGCTCAGATCCATGAAGGCTTCAGGAATCTCCGTAACCTCAAC 241
QY 324 CAACCTGATTCTCAATTTGCAATTAACCTACTGTTAAAGCGTTTATTTTGTCTAGAGTTTA 383
Db 242 CAGCCAGCAGCCAGCTCCAGCTGACCCAGCAATGGCCTGTTCTCAGCGAGGCGCTG 301
QY 384 AAATGTTGACAAATTTCCCTAGAGAGCTCAGAACTATATCATAGTAGGCTTTTACC 443
Db 302 AAGCTAGTGGATAAGTTTTTGGAGGATGTTTAAAAAGTTGTACCACTCAGAAAGCTTCACT 361
QY 444 GTTAATTTTGGTATCTAGGAGCTAAAAAGCAATTAATGATTATTTGAGAAAGGC 503

AGENCOURT_10034701 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483305
5', mRNA sequence.
BQ958958
BQ958958.1 GI:22374436
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2667 row: k column: 18
High quality sequence stop: 586.
Location/Qualifiers
1. .924
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6483305"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 251 a 253 c 226 g 193 t 1 others
ORIGIN
Query Match 18.6%; Score 283.8; DB 14; Length 924;
Best Local Similarity 58.4%; Pred. No. 4e-63;
Matches 513; Conservative 0; Mismatches 363; Indels 2; Gaps 1;
QY 84 AAAATTACTCCAAATTTAGCGGAATTTGCTTTCTTTGTATAGACAATTAGCTCATCAA 143
Db 2 AGATACCCCAACCTGGCTGAGTTGCGCTTCAGCCTATACCGCAGCTGGCAGACACAG 61
QY 144 AGTAATTCTACTAACATTTTTTTTAGTCTGTTTCTATTGCGCACTGCTTTGCGCATGTTG 203
Db 62 TCCACAGCAGCAATACTTCTTCTCCCACTGAGCATCGCTACAGCCTTTGCAATGCTC 121
QY 204 AGTTTAGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGTTTAACTTTAATTG 263
Db 122 TCCCTGGGGACCAAGGCTGACACTCAGATGAAATCCTGGAGGCGCTGAATTTCAACCTC 181
QY 264 ACCGAATCCCAAGACCCAAATTCAGAGGGTTTTCAAGAGTTGTTGAGAATTTGAAT 323
Db 182 ACGGAGATTCGGAGGCTCAGATCCATGAAGGCTTCAGGAATCTCCGTAACCTCAAC 241
QY 324 CAACCTGATTCTCAATTTGCAATTAACCTACTGTTAAAGCGTTTATTTTGTCTAGAGTTTA 383
Db 242 CAGCCAGCAGCCAGCTCCAGCTGACCCAGCAATGGCCTGTTCTCAGCGAGGCGCTG 301
QY 384 AAATGTTGACAAATTTCCCTAGAGAGCTCAGAACTATATCATAGTAGGCTTTTACC 443
Db 302 AAGCTAGTGGATAAGTTTTTGGAGGATGTTTAAAAAGTTGTACCACTCAGAAAGCTTCACT 361
QY 444 GTTAATTTTGGTATCTAGGAGCTAAAAAGCAATTAATGATTATTTGAGAAAGGC 503

Db 362 GTCAACTTCGGGACACCGAAGAGGCCAAGAACAGATCAACGATTACGTGGAGAAGGCT 421
 QY 504 ACCAGGGTAAGATCTTCACCTAGTTAAAGATTAAGATAGATCGTGATACCGCTTCGCCACTA 563
 Db 422 ACTCAAGGGAAATGTGGATTGGTCAAGGAGCTTTGACAGAGACACAGTTTTCCTCTG 481
 QY 564 GTTAACATATATTTTTCAGGCTAAGTGGGAACGTCCTTCGAGGTTAAAGATACTGAA 623
 Db 482 GTGAATTACATCTCTTTAAAGGCAATGGGAGAGACCCCTTTGAAGTCAAGGACACCGAG 541
 QY 624 GAGGAAGATTTTCATGTTGATCAAGTTACTACTGTCAAGATTTCCAAATGATGAAGAAAGACTG 683
 Db 542 GAAGAGGACTTCCACGTGGACCGAGCGACCCGTTGAAGTGCCTATGATGAAGCTTTA 601
 QY 684 GGTATCTTCAATATTCACCAATTCGAAATTAAGTTCTTGGGTCTTATTAAGTAT 743
 Db 602 GGCATGTTTAAATCAGCAGCTGTAGAAAGCTGTCCAGCTGGGTGCTGATGAATATAC 661
 QY 744 TTAGTAAAGCTACTGCTATTTTTCACGAGGTAAGCTTCAACATTTAGAG 803
 Db 662 CTGGCAATGCCACGCCATCTTCTCGCTGATGAGGGAACACTACAGCACCTGGAA 721
 QY 804 AATGAGTTCACATCATGACATTAATTAATTT-TTTTAGAGAACGAGGATCGTCGACG 861
 Db 722 AATGAACCTCACCCAGATATCATCNCAGTTCCCTGGAAATGAAGACAGAAAGCTCGT 781
 QY 862 CTTCCTGCACTGCCAAAGTTTAAGTATCACCGGTACTTACGACTTAAATCTGTTTAG 921
 Db 782 CCACCTTAGATTTCCCAACTGTCATTTACTGGAACATATGATCGAAGAGCGTCTCGG 841
 QY 922 GCCAGTTAGGTATTACCAAGTTTTTTTCTAACGGGTGCC 959
 Db 842 GTCAACTGGGCATCACTAAGGCTTTCAGCAATGGGGGC 879

RESULT 11
 BM924813 1194 bp mRNA linear EST 12-MAR-2002
 LOCUS AGENCOURT_6653921 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5761267
 DEFINITION 5', mRNA sequence.
 ACCESSION BM924813
 VERSION BM924813.1 GI:19375192
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1194)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL12809 row: j column: 20
 High quality sequence stop: 662.
 Location/Qualifiers
 1..1194
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone IMAGE:5761267
 /clone_lib="NIH_MGC_116"
 /lab_host="DH10B"
 /note="Organ: pooled colon, kidney, stomach; Vector:
 pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 3 colons, age 26 yo male, 49 yo
 female, 71 yo male colon; 46 yo male kidney, and pool of 2

FEATURES

source

RESULT 12
 BQ650189
 LOCUS
 DEFINITION
 ACCESSION

BQ650189 907 bp mRNA linear EST 15-JUL-2002
 AGENCOURT_8298326 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269613
 5', mRNA sequence.
 BQ650189

stomachs, 62 yo male and 70 yo female. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.4 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 023. Note: this is a NIH_MGC Library."

BASE COUNT 321 a 303 c 328 g 242 t
 ORIGIN

Query Match 18.0%; Score 274.8; DB 14; Length 1194;
 Best Local Similarity 61.2%; Pred. No. 9.7e-61;
 Matches 444; Conservative 0; Mismatches 282; Indels 0; Gaps 0;
 QY 448 ATTTTGGTGATCTGAGGAAGCTAAAAGCAAAATTAATGATTAATGTGAGAAAGGCACCC 507
 Db 7 AATTTCCGGGATCCGAAGAGGCCAAGAACAGATCAACGATTTAGCTGGAGAAAGGTTACT 66
 QY 508 AGGCTAAGATCTGTGACCTAGTTTAAAGANNTAGATCTGATACCGTCTTCGCACACTAGTTA 567
 Db 67 AAGGAAAATTTGGATTTGGTCAAGGAGCTTTGACAGAGACACAGTTTTCCTCTGGTGA 126
 QY 568 ACTATATTTTTTCAAGGGTAAAGTGGGAAGCTCTTTTCGAGGTTAAAGATACTGAAGAGG 627
 Db 127 ATTACATCTCTTTTAAAGGCAATGGGAGAGACCCCTTTGAAGTCAAGGACACACGGAAG 186
 QY 628 AAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTTTCCAAATGATGAAAGACTGGGTA 687
 Db 187 AGGACTTCCAGCTGGAGCAGGTGACCACCGTGAAGGTGCCTATGATGAAGCGTTTAGGCA 246
 QY 688 TGTTCAATATTTCAACATTTGCAAAAAATTAAGTTCTTCGGGTCTTATTAATCAAGTATTAG 747
 Db 247 TGTTTAACTCCAGCACTGTAAAGAGCTGCCAGCTGGGTGCTGCTGATGAATATCTCG 306
 QY 748 GTAAAGCTACTGCTATTTTTTTTTTACCAGCAAGGTAAGCTTCAACATTTTAGAAGATG 807
 Db 307 GCAATGCCACCGCCATCTTCTCTGCTGATGAGGGGAAACTACAGCACCTGGAATG 366
 QY 808 AGTTGACTCATGACATTAATTAATTTTACTAAATTTTACAGAGAGGAGATCGTCGAGCGTTCTC 867
 Db 367 AACTCACCCACGATATCATCAACAAAGTTCTCGAAATTAAGACAGAAAGGTCTGCCAGCT 426
 QY 868 TGCACTTGCACAAAGTTAAGTATCACCGTACTTACGACTTAAATCTGTTTTAGGCCAGT 927
 Db 427 TACATTTACCCAAACTCTCCATTAAGTATGATCTGAAGAGCGTCTCGGGTCAAC 486
 QY 928 TAGGTATTTACCAAGTTTCTTAACGGTCCCGATTTGAGTGTGTTACTGAAGAAGCTC 987
 Db 487 TGGCATCACTAAGTCTTCAGCAATGGGCTGACCTCTCCGGGGTCCACAGAGAGGACAC 546
 QY 988 CATTAATTTGAGTAAAGCTGTTCAACAAAGCCGCTTAACATTTGATGAAGAGGGTACCG 1047
 Db 547 CCCTGAAGCTCTCCAGGCGCTGCATAAGCTGTGCTGACCATCGACGAGAAAGGACTG 606
 QY 1048 AGGCGCGCGGCGGTATGTTCTCTGGAAGCTATTCCCAATGAGCATTTCCACCAGAAAGTTAAAT 1107
 Db 607 AAGCTGCTGGGCCATGCTTTTGAAGGCCATACCCATGCTATCCCCCGGAGGTCAGT 666
 QY 1108 TTAATAAACCATTCGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1167
 Db 667 TCAACAAACCCCTTTGTTCTTAAATGATGAACAAAATACCAAGTCTCCCTCTCTCATGG 736
 QY 1168 GTAAGG 1173
 Db 727 GAAAG 732

```

VERSION      B0650189.1  GI:21774361
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 907)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgabbs-remail.nih.gov
             Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM2443 row: g column: 22
             High quality sequence stop: 650.
FEATURES     Location/Qualifiers
             1..907
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:6269613"
             /clone_lib="NIH_MGC_100"
             /tissue_type="hepatocellular carcinoma, cell line"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: liver; Vector: pOFB7; Site_1: XhoI; Site_2:
             EcoRI; cDNA made by oligo-dT priming. Directionally cloned
             into EcoRI/XhoI sites using the following 5' adaptor:
             GGCACGAG(G). Size-selected >500bp for average insert size
             1.8kb. Library constructed by Ling Hong in the laboratory
             of Gerald M. Rubin (University of California, Berkeley)
             using ZAP-cDNA synthesis kit (Stratagene) and Superscript
             II RT (Life Technologies). Note: this is a NIH_MGC
             Library."
BASE COUNT   245 a 237 c 226 g 198 t 1 others
ORIGIN
Query Match 17.98; Score 273.4; DB 14; Length 907;
Best Local Similarity 59.48; Pred. No. 2.1e-60;
Matches 498; Conservative 0; Mismatches 337; Indels 3; Gaps 2;

QY 336 CAATTCGCAATTAACCTACTGGAACGGTTTATTTTGTCTGAAGTTTAAATGGTTGAC 395
    || || || || || || || || || || || || || || || || || || || ||
DB 7 CAGCTCCAGCTGACCCACCGCAATGGCTGTCTCTCAGCGAGGCGCTGAAGCTAGTGGAT 66
QY 396 AAATTCCTAGAGAGCTCAAGAACTATATCATAGTAGGCGTTTACGGTTAATTTTGGT 455
    || || || || || || || || || || || || || || || || || || || ||
DB 67 AAGTTTTTGGAGGATGTTAAAAAGTTGTACCACTCAGAAGCGCTTCACTGTCAACTCGGG 126
QY 456 GATACCTAGGAGCTPAAAAAGCAATTAATCATATGTTGAGAAAGGCACCGGGTAAG 515
    || || || || || || || || || || || || || || || || || || || ||
DB 127 GACACCGAAGAGGCCAAGAAACAGATCAACATTACGTGGAGAGGGTACTCAAGGGAA 186
QY 516 ATCTGTTGACCTAGTTAAAGAATTAGATCGTGATACCGTCTTCGCAGTATTAATATT 575
    || || || || || || || || || || || || || || || || || || || ||
DB 187 ATTGTGGATTGGTCAAGAGCTTCACAGAGACACAGTTTTTGTCTGTGTAATTACATC 246
QY 576 TTTTTCAGGGTAAGTGGAGAGCTCCTTTCGAGGTTAAAGATCTGAAGAGAGAGATTTT 635
    || || || || || || || || || || || || || || || || || || || ||
DB 247 TTTCTTAAAGGCAATGGGAGAGACCCCTTTGAAGTCAAGGACACCGAGAGAGGACTTC 306
QY 636 CATGTTGATCAAGTTACTACTGTCAAGTCCCAATGATGAAAGACTGGGTATGTTCAAT 695
    || || || || || || || || || || || || || || || || || || || ||
DB 307 CAGGTGGACCAAGTACCAACCGGTGAAGGTGCTATGATGAACGCTTTAGGATGTTTAC 366
QY 696 ATTCAACATTTGCAAAAAATTAAGTTCTTTGGGTCTTTTAAATCAAGTATTTAGGTAACGCT 755
    || || || || || || || || || || || || || || || || || || || ||
DB 367 ATCCAGCACTGTAAAGAGCTGTCACAGCTGGGTGCTGCTGATGAATACTCGGCAATGCC 426

```

```

QY 756 ACTGCTATTTTTTTTACCAGACGAAGGTAGCTTCAACATTTAGAGAATGAGTTGACT 815
    || || || || || || || || || || || || || || || || || || || ||
DB 427 ACCGCCATCTTCTCTGCTGATGAGGGGAACTACAGCACCTGGGAAATGAACCTACC 486
QY 816 CATGACATTAATTAATAATTTTATAGAGAACGAGGATCGTCGTAGCGCTTCTCTGCACCTG 875
    || || || || || || || || || || || || || || || || || || || ||
DB 487 CACGATATCATCACCAGTCTCTGGAATATGAAGACAGAGAGTCTGCCAGCTTACATTTA 546
QY 876 CCAAAAGTTAAGTATCACCGGTACTTACGACTTTAAATCTGTTTATAGGCGAGTTAGTATT 935
    || || || || || || || || || || || || || || || || || || || ||
DB 547 CCAAAACTCTCCATTAATCTGGAACCTATATCTGAAGAGCGTCTCGGTCAACTGGGCATC 606
QY 936 ACCAAAGTTTCTTAACGCTGCCGATTTGAGTGTGTTACTGAAGAAGCTCCATTAATA 995
    || || || || || || || || || || || || || || || || || || || ||
DB 607 ACTAAGGTCTTCAGCAATGGGCTGACCTCTCCGNGTCCAGAGAGAGCCCTCGAAG 666
QY 996 TTGAGTAAAGCTGTTTACAAAGCCGCTTCTTAATGATGATAAAGGGTACCGAGCCGCC 1055
    || || || || || || || || || || || || || || || || || || || ||
DB 667 CTCTCAAGGCCGTGCATAAGGCTGTGCTGACCATCGACGAGAGAAAGG-ACTGAAACTGCT 725
QY 1056 GGGGCTATGTTCTTGAAGCTATTTCAATGAGCATTCACAGAGAGTTAAATTTAA--TA 1113
    || || || || || || || || || || || || || || || || || || || ||
DB 726 GGAGCCATGTTTTTAGAGGGCTTACCCATGTCTATCCCCCCCCGAGTCAAGGTTCCACAA 785
QY 1114 AACCATTCGTTTTTCTGATGATCGAGCAGACACATAAAAGCCCATTTGTTTATGGGTAA 1171
    || || || || || || || || || || || || || || || || || || || ||
DB 786 ACCCTTTGCTCTTTTAATGATGGACCAATACCAAGTCTCCCTCTCTCCGGGGAA 843

RESULT 13
LOCUS      BQ648524
DEFINITION AGENCOURT_8297828 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269919
            5', mRNA sequence.
ACCESSION  BQ648524
VERSION    BQ648524.1  GI:21772696
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 959)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2444 row: d column: 16
            High quality sequence stop: 674.
FEATURES   Location/Qualifiers
            1..959
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:6269919"
            /clone_lib="NIH_MGC_100"
            /tissue_type="hepatocellular carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: liver; Vector: pOFB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-cDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH_MGC
            Library."

```

VERSION BQ646948.1 GI:21771120
KEYWORDS EST.
SOURCE human.

| | |
|----|-----|
| QY | 432 |
| DB | 508 |

AGGCTTTACCGTT
| | | | |
AAGCCTTCACTGT

AGGCTTTTACCGTTAATTTGGTGATACTGAGGAAG

AGGCTTTTACCGTTAATTTGGTGATACTGAGGAAGCTAAAAAGCAAATTAAATGATTAT 491
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAGCTTCACCTGTCAACTTCGGGGCACCAGAACAGGCCACGCACACATCACCATTAAC 567

